

inc consectetur hendrerit. Duis laoreet, nisi in iaculis consequat, ante  
at felis. Quisque varius neque ut nisi. Aenean et arcu sed risus phare  
smod quis, velit. Pellentesque lacus dolor, ullamcorper id, ornare p  
dolor elit, dignissim at, mollis sed  $A^2 + B^2$ ; nulla. Nam nisi augue, fauc  
vitae, eleifend id, cursus at, m  $A^2 + B^2$ ; nulla. Nam nisi augue, fauc  
ctetur, nisi aliquet tincidunt  $A^2 + B^2$ ; nulla. Nam nisi augue, fauc  
entum sodales. Pellentesque  $A^2 + B^2$ ; nulla. Nam nisi augue, fauc  
illus at eros, Nunc purus sapien  $A^2 + B^2$ ; nulla. Nam nisi augue, fauc  
convallis fringilla. Nunc portit  $A^2 + B^2$ ; nulla. Nam nisi augue, fauc  
tincidunt tellus neque eu lectu  $A^2 + B^2$ ; nulla. Nam nisi augue, fauc  
nectus et netus et malesuar  $A^2 + B^2$ ; nulla. Nam nisi augue, fauc  
ger in eros. Aliquam dar  $A^2 + B^2$ ; nulla. Nam nisi augue, fauc  
ulvinar quis, tincidunt  $A^2 + B^2$ ; nulla. Nam nisi augue, fauc  
Phasellus vel nisi vita  $A^2 + B^2$ ; nulla. Nam nisi augue, fauc  
elit, egestas id, ege  $A^2 + B^2$ ; nulla. Nam nisi augue, fauc  
s nec mauris. Nam  $A^2 + B^2$ ; nulla. Nam nisi augue, fauc  
ligula quis lorem,  $A^2 + B^2$ ; nulla. Nam nisi augue, fauc  
quat venenatis. V  $A^2 + B^2$ ; nulla. Nam nisi augue, fauc  
blandit non, felis  $A^2 + B^2$ ; nulla. Nam nisi augue, fauc  
s. Pellentesque h  $A^2 + B^2$ ; nulla. Nam nisi augue, fauc  
llis tincidunt. In tortor. Intege  $A^2 + B^2$ ; nulla. Nam nisi augue, fauc  
is, egestas nec, ligula. Mauris  $A^2 + B^2$ ; nulla. Nam nisi augue, fauc  
is in enim id ligula sodales  $A^2 + B^2$ ; nulla. Nam nisi augue, fauc  
pretium posuere. Class apte  $A^2 + B^2$ ; nulla. Nam nisi augue, fauc  
quis, commodo sit amet, ve  $A^2 + B^2$ ; nulla. Nam nisi augue, fauc  
. Aenean semper diam sit ar  $A^2 + B^2$ ; nulla. Nam nisi augue, fauc  
olestie quis, vestibulum a, t  $A^2 + B^2$ ; nulla. Nam nisi augue, fauc  
massa nisi, rhoncus nec, por  $A^2 + B^2$ ; nulla. Nam nisi augue, fauc  
lit id tortor. Duis adipiscin  $A^2 + B^2$ ; nulla. Nam nisi augue, fauc  
erdiet erat. Maecenas lectu  $A^2 + B^2$ ; nulla. Nam nisi augue, fauc  
ue porta. Proin vitae erat. Cl  $A^2 + B^2$ ; nulla. Nam nisi augue, fauc  
parturient montes, nascetur  $A^2 + B^2$ ; nulla. Nam nisi augue, fauc  
get nisi. Morbi feugiat. Praes  $A^2 + B^2$ ; nulla. Nam nisi augue, fauc  
sent neque odio, egestas quis,  $A^2 + B^2$ ; nulla. Nam nisi augue, fauc  
t mauris. Nam pellentesque  $A^2 + B^2$ ; nulla. Nam nisi augue, fauc

Social Sciences and Humanities

# Roadmap Working Group Report 2008



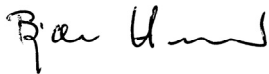
European Strategy Forum  
on Research Infrastructures

## Foreword

This is the 2008 report of the Social Sciences and Humanities Roadmap Working Group (SSH RWG) to the European Strategy Forum on Research Infrastructures (ESFRI).

The SSH RWG received one proposal for a new or upgraded pan-European Research Infrastructure (RI), submitted by the Lithuanian ESFRI delegate. A review of the proposal was undertaken based on a presentation by representatives of the RI to a RWG meeting, together with a pre-review carried-out in accordance with the ESFRI Terms of References, Procedures and additional rules. Conflicts of Interest were also declared by the RWG members and addressed accordingly.

In accordance with its mandate, this Report outlines the scientific landscape of the Social Sciences and Humanities, suggesting ways to follow-up the initiatives from the 2006 ESFRI Roadmap Report, together with a model for cooperation between those SSH RI's which received funding for a Preparatory Phase Project (PPP) and other SSH RI's.



Bjørn Henrichsen

Chair, SSH RWG.



# Contents

<b>1 Executive Summary .....</b>	<b>5</b>
Mandate and areas of interest .....	5
Projects from 2006 Roadmap Report .....	5
Proposals put forward to ESFRI for the Roadmap Update .....	7
The Landscape .....	7
Infrastructural System .....	9
To Sum Up: .....	10
<b>2 Methodology .....</b>	<b>11</b>
Conflict of Interests .....	11
Evaluation Methods for New Proposal .....	11
Evaluation Methods for projects on the 2006 Roadmap Report .....	11
<b>3 Scientific Landscape of Social Sciences and Humanities .....</b>	<b>12</b>
Technology of Data Access and Collection .....	12
Infrastructure needs .....	14
Long-Term Strategic Goals .....	19
<b>4 Towards an Infrastructural System .....</b>	<b>21</b>
Cooperation .....	22
Initial Phase .....	23
Need for Elaboration .....	24
A Cooperation Task Force .....	24
<b>5 Annexes .....</b>	<b>26</b>
Annex 1: Terms of Reference .....	26
Annex 2: Procedures and added criteria .....	29
Annex 3: Template use for reviewing the proposal .....	30
Annex 4: Assessment of the Received Proposal .....	31
Annex 5: Composition of Roadmap Working Group .....	33
Annex 6: Roadmap Working Group's meetings .....	34
Annex 7: Declarations of Conflicts of Interest .....	35
Annex 8: Proposal Received .....	36
Annex 9: References .....	41
Annex 10: Infrastructures put forward for the Roadmap Update .....	42
<b>CESSDA - Council of European Social Science Data Archives .....</b>	<b>42</b>
<b>CLARIN - Common Language Resources and Technology Infrastructure .....</b>	<b>43</b>
<b>DARIAH - Digital Research Infrastructure for the Arts and Humanities .....</b>	<b>44</b>
<b>ESS - European Social Survey Upgrade .....</b>	<b>45</b>
<b>SHARE –Upgrade of the Survey of Health, Aging and Retirement in Europe .....</b>	<b>46</b>



# 1 EXECUTIVE SUMMARY



## ***Mandate and areas of interest***

In this reporting period, the main tasks for the Social Sciences and Humanities Roadmap Working Group (SSH RWG) have been to:

- monitor and advise ESFRI on the optimal procedures in support of the implementation of the Roadmap;
- follow the general development of the research infrastructure landscape across the SSH, both within the ERA and globally;
- advise ESFRI on ways to stimulate the mature proposals on the 2006 ESFRI Roadmap;
- evaluate new proposals forwarded by ESFRI delegations.

The SSH RWG has not looked at more specific infrastructural needs in the different research fields within the Humanities and Social Sciences. This task was clearly outside the group's mandate and capacity given the time limitations imposed by the current phase of the ESFRI Roadmap process. Instead the RWG has aimed to provide broader and more general views in relation to the further development and advancement of RI's within both the Humanities and the Social Sciences.

## ***Projects from 2006 Roadmap Report***

### **Status and Perspective**

Five of the six SSH RI's included on the 2006 ESFRI Roadmap applied for and received funding from the European Commission for a Preparatory Phase Project, providing support from FP7 to establish the legal, organisational and financial framework for the construction and operation of the new or upgrading facility.

These five projects are the:

- Council of European Social Science Data Archives (CESSDA);
- Common Language Resources and technology Initiative (CLARIN).
- Digital Research Infrastructure for the Arts and Humanities (DARIAH);
- European Social Survey (ESS);
- Survey of Health, Ageing and Retirement in Europe (SHARE).

In addition, a further facility included on the Roadmap – the European Research



Observatory for the Humanities and Social sciences (EROHS) -- did not apply for funding under the FP7 PPP call. Based on an evaluation of the present status of EROHS, the SSH RWG recommends that EROHS is not included in the update of the ESFRI Roadmap. This proposal is also partly based on the suggestion from the SSH RWG to nurture and facilitate greater cooperation and interchange between the five existing SSH funded PPP projects and, in so doing, explore the potential for establishing a broader infrastructural system which could incorporate and support additional RIs and research networks within the SSH. In some respects such cooperation between mature, on-going facilities, together with the harnessing of other infrastructural developments within the SSH overlaps with the key *raison d'être* of the earlier EROHS proposal. However, in order to succeed in establishing an integrated system, based on cooperation and interchange between different RIs, a limited amount of funding would be needed.

#### Short description of the six "2006 projects"

*Council of European Social Science Data Archives (CESSDA)* is a distributed RI that provides and facilitates access to high quality data and supports researchers in the use of these data. The CESSDA RI consists of a network of national organisations currently extending across 21 countries throughout Europe. In total it holds, curates and provides access to over 20,000 separate data collections, supporting a European-wide user community of some 200,000 researchers and learners. CESSDA already operates within a global data environment, with reciprocal data access arrangements and agreements established with other data holding organisations worldwide.

A major upgrade aims to develop CESSDA from the current situation in which the member organisations work with limited national resources to create a common platform, sharing a common mission, to a fully-integrated data infrastructure for the SSH. Such an upgrade will allow the discovery of datasets and data-related materials in an integrated cross-national

environment.

The *Common Language Resources and technology Initiative (CLARIN)* is a large-scale pan-European coordinated infrastructure effort to make language resources and technology available and useful to scholars of all disciplines, in particular within the humanities and social sciences. It aims to provide a comprehensive and easily accessible archive of language resources and technology, covering not only the languages of all EU Member States, but also languages and language issues related to migration.

The tools and resources will be interoperable across languages and domains and will contribute towards addressing the issue of preserving and supporting multilingual and multicultural European heritage. It will be based on a virtual grid-type infrastructure employing Semantic Web technology.

The *Digital Research Infrastructure for the Arts and Humanities (DARIAH)* will be based upon an existing network of data centres and services, and aims at providing an infrastructure across the arts and humanities in order to access resources for researching the cultural heritage of Europe. It thereby contributes to a better common understanding of cultural diversity and its history within Europe. It will help to cross cultural boundaries and to create a new European coherence based on mutual understanding and on true integration of the uniquely rich European traditions. DARIAH will be based on partner organisations that have national infrastructures in place and are already collaborating.

The *European Research Observatory for the Humanities and Social sciences (EROHS)* is a proposal for a system that should operate both as a central and distributed facility with a strong physical hub working in close conjunction with a number of spokes across Europe, harnessing European expertise through a coordinated yet decentralised network. It should be organised to promote and ensure cooperation and integration of data, technologies and policies.



The *European Social Survey (ESS)* monitors long-term change in social values throughout Europe, producing data relevant to academic debate, policy analysis and better governance. ESS data are made publicly available via the internet as soon as they are available, with no prior “privileged” researcher access.

The required upgrade is to unify, regularise and secure the funding for the RI as a whole over an extended period, though naturally with periodic reviews. A large and complex time series such as the ESS requires continuity of funding as a prerequisite of appropriate planning.

The *Survey of Health, Ageing and Retirement in Europe (SHARE)* provides a data infrastructure for fact-based economic and social science analyses of the on-going changes within Europe due to population ageing. Preliminary collection of data started in 2002, and in 2004 a first wave of data on the economic, health and family conditions of about 27,000 respondents aged 50 and over was collected in 11 European countries. The participating countries covered all EU15 regions. The data are harmonised to facilitate cross-national comparative research. The second wave of data collection is currently on-going and has been extended to include Poland, the Czech Republic and Ireland. A third wave of data collection will specialise in life histories.

### **Proposals put forward to ESFRI for the Roadmap Update**

The SSH RWG received only one new proposal – the *European Election Studies (EES)* -- submitted by the Lithuanian ESFRI delegation.

The EES is an infrastructure for research into citizenship, political participation and electoral democracy throughout the European Union. It is used to address scientific questions regarding representation and democracy across the ERA. The user community extends to every country in the EU, in the first place to researchers in European studies,

political economy, political science and social psychologists, but also to users outside the scholarly community.

The proposal is basically to continue support for a database which has been in existence since 1979, being upgraded after each round of European elections. Twenty-four European countries contributed financially to the 2004 study.

The EES proposal clearly states that it is not expecting to receive funding via ESFRI at this time. The European Commission has currently secured funding for the EES till 2012 and an upgrade of the RI is already underway based to a large extent on this funding. Instead, EES wishes to claim recognition from ESFRI as a mature research infrastructure, and to be placed on the ESFRI Roadmap.

An internal group selected from among the SSH RWG membership reviewed the proposal. This review was based on the proposal from Lithuania, a presentation by Profs. Algis Krupavicius and Cees van der Eijk to a RWG meeting in January, and followed the ESFRI Terms of References for the review of project proposals (including procedures and additional rules, see Annex 3). At its meeting in April, the SSH RWG assessed the EES proposal based on the various inputs mentioned above and concluded that it was not appropriate for the EES to be included within the 2008 Roadmap update.

### **The Landscape**

#### **Technology**

The practice of the Social Sciences and Humanities has been transformed with the emergence of new survey and information technology. This transformation demands new kinds of interactions which require closer collaborations between the researchers, more sophisticated data infrastructures and the establishment of state-of-the-art software tools.



### Infrastructure needs

The Humanities and Social Sciences need access to a broad range of established data collections and archiving. It is also important that the collections are built-up and made available. Data collections of different types are an essential element of research infrastructural needs within the two domains. Although the nature of the information required differs both between and within the two, the infrastructural requirements of the two share similar characteristics. However, it should be noted that in the case of the Humanities there is a particular problem due to the immaturity (from a technical standpoint) of many RIs and that the landscape is incredibly fragmented. Beyond the collection of data (and retrospective digitisation of data in the Humanities) it is essential to secure free (at the point of use) and easy access to public funded data resources. In line with OECD recommendations, public-funded data, which include, for instance, data collected by public-funded research teams as well as official administrative data and surveys, should be defined as public property and a resource that should be publicly-shared for common benefit. They should be made freely-available and barriers to access should be restricted to ensuring and maintain the data subject's confidentiality and upholding intellectual property rights.

Preservation of public-funded research data for long-term scientific re-use through formal archiving is equally important. Scientific activities, both within the Humanities and the Social Sciences, not only use but can also produce data (research data). Such data can also be shown to be of important value for secondary analysis – facilitating comparative research, informing follow-up studies and developing methodology.

Given the technological possibilities, the Humanities have a demand for open, online access to a wide range of primary source materials housed in repositories and digital archives. These technological possibilities are not only interesting for academic communities, but also for planning agencies and commercial players.

One major objective will be to provide seamless access to data across repositories, nations and research purposes. To encourage data sharing and interoperability across communities and software systems, standardization of metadata and data must be a high priority and should be encouraged in future projects and initiatives.

One of the major challenges in the building of new European research infrastructures lies in the interoperability between different domains of research. Currently, research infrastructures are not only separated by national borders, but also by disciplinary fields. This is a particular problem for the Humanities and Social Sciences.

One of the main tasks of the new European RIs must therefore be to leave the current distinctions and divisions behind and create research tools and instruments that can be readily accessed and used across all domains.

In the Social Sciences it is of particular importance to support a unified approach to social and behavioural science. Equally, in the Humanities, given that many of the RI initiatives are at an earlier stage of development and maturity it is essential to foster the coordination of effort in order to maximise interoperability and impact.

In most European countries data are often preserved and made available by national repositories and digital archives. It is of crucial importance that these data resources can be discovered and made easily available for international use through integrated internet-based services, employing emerging grid technologies as appropriate.

As a single example to illustrate this goal, technologies meeting a subset of these requirements have been developed by EC-funded projects such as the MADIERA project (<http://www.madiera.net/>). The Council of European Social Science Data Archives (CESSDA) is now strengthening its European data location and access services, building on the technologies developed through the MADIERA and earlier EC-funded projects.



Within the Social Sciences, one major obstacle to the access of official empirical statistical data across Europe is the multitude of data access policies and regulations implemented by national governments. In order to make it easier to bring together data from various countries a broad mapping of all present rules and requirements is needed, followed by harmonization and a single transferable 'one-stop' access approach as far as is possible.

Also within the Humanities the CLARIN project represents similar ideas by aiming to establish an integrated and interoperable RI of language resources and its technology. The project, which originally is developed from three separate initiatives, aims to lift the current fragmentation, offering a stable, persistent, accessible and extendable infrastructure by incorporating, and contribute to, Semantic Web technology. It includes Data Grid technology to connect the repositories and incorporates multi-lingual language processing technology that supports cultural and linguistic integration.

#### Long-term Strategic Goals for the Infrastructures

These goals can be formulated as:

- increasing the availability of quality comparative data and enhancing data modeling;
- the development of data integration and language tools;
- coordination of and facilitating access to research resources and materials through integrated systems and processes enabling;
- the digitization, enhancement and repurposing of key analogue data resources.

The first of these strategic goals is linked to the need for European-wide data; the second encompasses the interoperability of data and languages while the third is directed towards the harmonization of data access policies, as well as creating greater interoperability

between existing and emerging research infrastructures across the Social Sciences and the Humanities in general. The last, is particularly relevant to the Humanities, recognising the special and large scale problem which exists within these disciplines which impacts both on resource discovery and access, as well as limiting methodological development.

### *Infrastructural System*

Cooperation and collaboration between both existing and emerging research infrastructures, together with related research networks and centres of expertise should be supported as a priority. This is viewed as being particularly important within the social sciences and humanities where (in comparison to certain other discipline areas) the strength of the research infrastructure environment lies in the collective or aggregate rather than the individual. In order to facilitate greater cooperation and collaboration between the different infrastructure operators, explore ways in which synergies between them can be harnessed and foster integration with other related research networks (new and established) the SSH RWG recommends that a new so-called Task Force<sup>1</sup> be established.

This new body will both develop and augment the work of the existing RWG in order to make recommendations and proposals in support of the goal of ensuring that the European SSH researcher community has access to a common integrated network of RIs. Of particular importance will be the need to explore and provide potential solutions to the problem of multiple (and non-exchangeable) access and authentication systems, differential employment of metadata

---

<sup>1</sup> Jane Ohlmeyer, Ireland, and Gert Wagner, Germany, did not support the proposal to form a Task Force



standards and operation of various rights management frameworks. Creating greater standardisation and procedural harmonisation will thereby add value to the investments made in individual structures and make a significant contribution to ERA in relation to research activities within the Social Sciences and Humanities.

The Task Force should consist of two separate bodies that are based on participation by researchers and academic communities: (1) a Steering Board and; (2) a Scientific Advisory Board (SAB). Members of both groups should be appointed on the basis of consultations with European Science Foundation, ESFRI and the countries that are participating in the Task Force. The research infrastructures that will participate actively in the cooperation must be represented in at least one of the two bodies. The rules for appointments need, however, to be discussed in detail.

The Task Force will need a 'neutral' secretariat located outside the involved infrastructures that should serve both the Steering Group and the SAB.

Different models of funding should be looked into as part of the pilot project, and could include funding from the EC, funding from national Ministries, agencies, or Research Council; or a mixed model.

A Task Force that will contribute to a better coordination of infrastructural initiatives will be an important tool in securing and maximizing the current and future investments within the two fields.

What is necessary in the first phase is:

- the adoption of a set of optimal and consistent standards and practices that are sine qua non for public research;
- an assessment of what has to be done to make existing infrastructures compatible with these standards and practices.

### *To Sum Up:*

The SSH RWG recommends that

- Five out of the six SSH projects in the 2006 ESFRI Roadmap should be listed also in the 2008 Update. These are:
  - CESSDA
  - CLARIN
  - DARIAH
  - ESS
  - SHARE.
- EROHS should be excluded from the 2008 Roadmap Update.
- The European Election Study (EES) should not be included within the 2008 Roadmap Update
- A coordinating Task Force consisting of two bodies; Steering Group and Scientific Advisory Board, should be established following a pilot study. The Task Force should be served by a secretariat



## 2 METHODOLOGY

### **Conflict of Interests**

Following the Terms of Reference (Annex 1) the members of the RWG declared potential conflicts of interest whenever they occurred in relation to the specific projects being discussed. These potential conflicts did not necessarily bar them from participating in the RWG's discussions (from "Terms of Reference of Roadmap Working Groups"). For an overview see Annex 4.

### **Evaluation Methods for New Proposal**

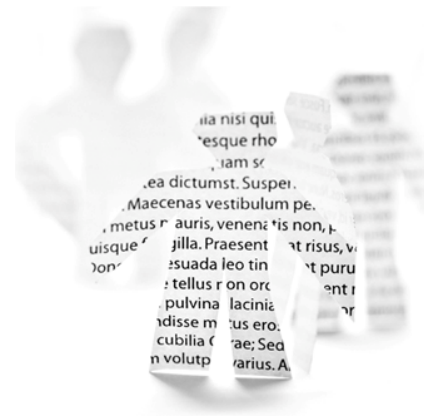
The Roadmap Working Group received only one proposal; the *European Election Studies* (EES), submitted by the Lithuanian delegation to ESFRI. The project was presented at a SSH RWG meeting.

An internal group was selected from among the Roadmap Working Group membership and reviewed the proposal. The Group consisted of:

- Karl Müller, Austria;
- Jan Karel Koppen, The Netherlands;
- Dominique Joye, Switzerland.

The review was carried out using a specially constructed template (Annex 4).

The RWG's recommendation is based on the proposal from Lithuania, the presentation by EES' representatives (Algis Krupavicius and Cees van der Eijk) to a RWG meeting, and the pre-review done in accordance with the ESFRI Terms of Reference, including



Procedures and added rules (see Annex 2).

### **Evaluation Methods for projects on the 2006 Roadmap Report**

The RWG has followed closely the status and development of the projects on the 2006 Roadmap Report. Five of the projects applied for and received EC funding (FP7) for a Preparatory Phase Project (PPP).

The SSH RWG has seen cooperation between the SSH Projects in the 2006 Roadmap as one of its main activities. PPP coordinators were therefore invited to a meeting of the RWG. A special meeting with the five project coordinators and the Chair and Secretary of the SSH RWG was also arranged.

It is agreed that it is in the interest of all the PPPs to try to secure close cooperation between them. The SSH RWG sees such cooperation as an essential part of establishing a new infrastructural system for the SSH-area. Based on this cooperation, it is the hope of the SSH RWG that other research infrastructures (included in later ESFRI Roadmaps as well as others) will later be able to join an integrated network based on agreed and shared principles for cooperation.



### 3 SCIENTIFIC LANDSCAPE OF SOCIAL SCIENCES AND HUMANITIES



#### *Technology of Data Access and Collection*

The practice of the Social Sciences and Humanities has slowly, but profoundly, been transformed through the emergence of new information technology. Digital resources, computer networks, and software tools influence to a great extent the nature of the human record, the way it is understood and the ways in which these interpretations are communicated.

The development of new technologies for access to data and enhanced capacities demand new kinds of interactions which transect and cut across research fields, institutions and national boundaries. Advanced research within the Humanities and Social Sciences will most probably increasingly involve and require international collaborations, sophisticated data infrastructures, and state-of-the-art software tools for publishing, documenting, finding, accessing and analyzing data online regardless of location or institutional affiliation. These new technologies will enable automating research and collaboration.

However, the development of tools for the Social Sciences and Humanities presents major challenges regarding access to resources, particularly in terms of the development of appropriate software tools and application of metadata standards and protocols which are needed to fully utilize the vast amount of data and information potentially available. The Social Sciences and Humanities face substantial problems in

developing tools that may not be commercially viable, and thus need levels of support and robustness that are normally associated with commercial partnerships.

Digitizing the products of human culture and society pose intrinsic problems of complexity and scale. The complexity of the record of human cultures — a record that is multilingual, historically specific, geographically dispersed, and often highly ambiguous in meaning — makes digitisation complex and the process of implementation difficult and expensive.

A major issue - representing considerable problems especially of a financial nature - for the Humanities and to a lesser extent the Social Sciences is how to convert the immense corpora and data, whether in text or other analogous formats, from the physical into digital form, or create digital representations of physical objects in libraries, archives, and in museums. Compared to this, recycling of the digital-born material will not be a huge problem in the decade to come. Retro-digitisation of critical masses of research relevant material on a systematic basis is of the utmost importance both to e-science and e-learning in the coming years.

In launching the “i2010: Digital Libraries initiative” in 2005, the EU declared that the cultural heritage of the European countries should be digitized by all countries in order to make the foundations of a new digital



infrastructure – the so-called *European Digital Library*. Now officially renamed *Europeana*, it remains a vast project aimed at providing a one-stop portal comprising integrated search facilities to all digital cultural heritage objects in Europe in by 2009-10. It is based on a cooperation of all types of European archives, research libraries, and museums, gathered in a consortium under the leadership of CENL, The Conference of European National Librarians, comprising 9 organizations within the library, archive, museum, and audio-visual institutional sectors of Europe. The aim is to comprise more than 12 million digital documents by 2012, a goal that will easily be fulfilled, but not without severe problems for the research communities, as mass digitisation of books, journals, newspapers, and archival collections has hardly begun. Despite the problems of relevant and adequate content, the network of European national and university libraries, national archives, including regional archives and museums are now building up a digital infrastructure of relevance to all subjects within the Humanities and Social Sciences. This is based on the division of tasks and work, resource sharing and cooperation. All subject-based research infrastructures will have to take this parallel development into account in the future if they are to avoid duplication of effort and services.

The present major challenge is now to create pan-European infrastructural systems that are needed by the Social Sciences and Humanities to utilize the vast amount of data and information that already exists or are being generated throughout Europe. Today the Social Sciences and Humanities are, however, hampered by the fragmentation of the scientific information space and despite initiatives such as *Europeana*, the lack of digitized data in many areas. Data, information and knowledge are scattered in space and divided by language, cultural, economic, legal, and institutional barriers.

A similar situation is found in the case of comparative social science research. In consequence, too much research is based on data from a single nation, carried out by a single-nation team of researchers and communicated to a single-nation audience. This state of affairs is preventing the development of a comparative and

cumulative research process integrating the European Research Area.

Information for research purposes is, however, not a scarce resource in Europe, either for the Social Sciences or for the Humanities. Well-developed official statistical systems combined with a variety of academically driven data gathering programs and activities have and are producing a wealth of data and information about various aspects of European society. This also includes simulation systems and collections of multimedia content — images, text, moving images, and audio.

However, the majority of these resources are country or nation specific. They are produced to meet national requirements and collected by means of nation and language specific instruments based on local methodologies and classifications. They are normally documented only in national languages and rarely published for general use outside the country of origin. On top of this, nation-specific access restrictions will often prevent information to travel abroad. In addition, many data and information sources of importance in researching and understanding European society, economy and culture are created or compiled by commercial concerns and access and use by academics is effectively prohibited.

Yesterday's answers to these challenges would probably have been formulated in terms of centralisation and the establishment of large-scale European-wide institutions. Today's answers should rather focus on standardization, the power of emerging information and communication technologies, harmonization of data access restrictions and strengthening of and collaboration among already established groups and organisations engaged in the development of the European Research Area. Concerted efforts on a European scale are needed to bring about the necessary changes.

Access to and common exchange of data is a prerequisite for the fruitful utilisation of the possibilities offered to the Humanities and Social Sciences by the emerging technologies. There is a common international understanding emerging in which it is believed that access policies and



legal and financial conditions have a greater impact on the availability of data to the research community than do technological solutions. The challenge for the Social Sciences and Humanities, therefore, is more than fibre-optic cables, storage area networks or basic communication protocols.

It is important to stimulate social and scientific creativity. Emphasis on training of young scientific researchers is thus vital.

### **Infrastructure needs**

To encourage and nurture international collaborate research within the Social Sciences and Humanities in Europe, research infrastructures offering a series of crucial functions and resources must be established. Although they might very well be extensions of functions and resources already existing at a national level in some European countries. In some cases a model infrastructural service might even be present at a European level, but would need to be broadened in scope, strengthened or extended to other disciplines.

#### **Access to Data Collections**

Scientific databases are a crucial part of the pan-European infrastructures included in the 2006 ESFRI Roadmap Report, and more generally in the global science system. Effective access to these data, in a responsible and efficient manner, is required to take full advantage of the data and the possibilities offered by the rapidly evolving digital technology.

Accessibility to research data has become an important condition for maximising the research potential of new digital technologies and networks. An open and democratic access policy not only gives scientific advantages but it also provides greater returns from public investments in research activities.

The OECD's principle and guidelines for access to research data from public funding (passed in the OECD Council on December 14, 2006) give policy recommendations to governments and authorities on access to data produced and collected as a result of public funding. The principles and guidelines

are in line with the RWG's view of a necessary infrastructural development within the SSH area.

OECD's recommendations are based on the recognition that the contribution from the research community to the development of society is far more limited than the possibilities offered by technological development. This is despite vast public investments in data collections all over the World.

Access to and exchange of data and information are necessary for the research community to fully utilise emerging technological developments. However, presently access is to a large degree hampered by organisational, economical, legal and cultural conditions. These conditions potentially have a stronger impact on access and use of data than any technological solutions.

Similar views are expressed by The American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences. In their report *Our Cultural Commonwealth*, published in December 2006, they emphasise that the contributions to the development of knowledge and innovation by the research community are limited by conditions other than technological ones. Even if technology is a necessary building block in all research infrastructures, is it worthless unless other measures are included, at least if the objective is to support high quality international research.

Accordingly, the view of the SSH RWG's is that data access should embrace the following guarantees:

- that the research infrastructures are freely open to all research communities;
- that independence is maintained and that equal access is granted to all research communities;
- that both data and results generated by research activities are made available for further use and repurposing.
- Thus, the main conclusion is that access to public-funded data must be open, easy



and freely available. The only restrictions should be that access and long-term curation and preservation take full and due account of prevailing legislation concerning privacy, confidentiality, ethical considerations and the rights of both data subjects and data producers.

#### Data Quality

For both the Social Sciences and the Humanities, collection of data is essential and could be referred to as the basic fuel of the research process. Although the nature of the information needed may differ between the two domains, design and construction of infrastructures is fundamentally similar.

For the Social Sciences high quality cross-national data are essential for comparative research. Long-established data gathering projects like the Eurobarometers, ISSP and the World Value Studies have for decades been the only available data sources expressly designed to produce comparative empirical evidence.

The newly-established European Social Survey (ESS) has recently set new standards for this type of operation and demonstrated what can be achieved when data are collected according to the most stringent scientific methods. In this respect the ESS can serve as a model and example of best practice for other European-wide data gathering projects within other social science disciplines.

To facilitate this type of cross-national data gathering operation concerted effort are needed to standardise survey instruments, develop generic and multilingual survey modules to be used in a variety of projects, as well as to streamline cross-national sampling methodologies and related techniques.

Important data types are currently not as easily available as the research community will like to see. One example is contextual data. These are data that describe the environment in which individuals reside and behave and often provide a valuable supplement to survey data, for instance. By combining micro-level data with information

about the social context within the same analysis, it is possible to examine how external factors influence the attitudes and choices of individuals.

The availability of contextual data has improved to a certain degree over the last couple of decades. Researchers can now choose between numerous sources that provide potential contextual data on a variety of topics. Many of the data sources are even freely available on Internet. However, this positive assessment of the current situation of data availability must be qualified in five important respects:

- the technical solutions for disseminating data are sometimes burdensome and lack flexibility;
- there are still important areas where there is a lack of data;
- the coverage of some of the datasets is rather limited, both in temporal and geographical terms.;
- the quality and comparability of many widely used datasets in cross country comparisons have been challenged;
- the documentation of the data is often inadequate, making it difficult for users to examine the quality of the data.

There is clearly room for improvement. Briefly, two basic steps can and should be taken.

- Contextual data can be made more accessible by making high quality data available via a single portal. This would bring together data from different sources, and provide users with a more flexible way of accessing data.
- Problems of comparability and data quality must be better documented.

Concerted effort is needed to develop the methodologies to collect and standardise this type of data at a European level. Present-day digital societies are also producing a wealth of new data that until recently have infrequently been the object of scientific research. Examples of potential new data



resources include transactional data produced commercial concerns for business and customer profiling; CCTV footage; monitoring of internet use and other communication. Equally, digitisation of artefacts, including manuscripts and music, for example, offer researchers new ways of describing and analyzing art, sculptures, human languages, and behaviour.

#### Archiving of Research Data

Scientific activities, both within the Humanities and the Social Sciences, include not only users but also producers of data (research data). Archiving of public funded research data for scientific re-use is important in order to utilize in the best possible manner the resources public society has invested in research and data production. Equally important are the methodological and ethical arguments pointing out that archiving, data curation and thus access to the basis of scientific analysis is necessary for evaluation and insight in the results of research.

Archiving of research data is important for a number of reasons. The most central are to:

- ensure that important information about society is not lost for future generations, whether due to changes in the technology used to access or process the information, accidental loss, or decay of data storage media;
- permit re-use in future research projects, follow-up studies, time series and comparative analyses;
- avoid the collection or processing of identical information twice or more;
- avoid asking informants the same question twice or more (which undermines public motivation to respond to surveys and other studies);
- secure the possibility for future review and evaluation of research results.

A number of European countries as well as the European Commission are now focusing on archiving of research data and the demand for effective rules that regulate the delivery of publically-funded data by

researchers. In the case of several national research councils within Europe archiving of research data has for some time been a standard procedure and included in their research contracts.

Data are legally protected by EU data protection rules. Research projects which collect data are evaluated and are subjected to ethical review and data protection rules. The EC has, nevertheless, no clauses on archiving within their contracts. This policy is, however, now under discussion. Rules to secure that EC- funded data production will be archived for future access and re-use of data, is strongly recommended.

To sum up: neither individual researchers nor research projects should be allowed to define public-funded data as private property. Rather it should be viewed as a resource to be publicly shared for common benefit. The importance of maintaining managed access to data used in research projects should be emphasised both for reasons of transparency, accountability and explicability (e.g. quality evaluation of research results).

#### Digitalization

New technological possibilities lead to a public as well as scientific demand for open and online access to the full range of primary source materials housed in repositories such as museums, public research institutions, historical archives, local libraries and national research libraries, special collections, and privately held collections. This includes books and journals, newspapers and magazines, government and administrative documents, manuscripts, maps, photographs, satellite images, census data, recorded sound, film, and broadcast television.

New information technology offers ways to reunite disseminated collections, to compare exemplars, and to bring together the works of individual creators. The possibilities are unlimited. Today the internet gives access to billions of pages of information, the annual digital output totalling many times what most libraries in the world holds. Nevertheless a true network of cultural heritage information is far from being realised.



A phenomenon worthy of note is the fact that these technological possibilities are obviously interesting not only for the academic world, but also for commercial players like Google who, for instance, offer the possibility of searching the content of books and scholarly literature across a number of disciplines and resources.

In December 2004 Google™ announced that they would commence a massive digitisation programme of books – digitizing "the world's knowledge" (15 million books from six major research libraries originally, later expanded to more than twice this number) based on entire university library holdings from the 19th and 20th centuries, especially from the USA and UK (Cf. Ronald Milne: "The Google Mass Digitization Project at Oxford", *LIBER Quarterly*, vol. 16: 3-4, 2006).

This was extended to other European university and state libraries, more than 25 in all, in 2008. The resulting digital images should either be accessible in full text online via *Google Scholar*™ or searchable without access to the full text, according to the copyright clearance and rights management issues.

This project was considered an enormous challenge to almost all European countries, as it could be foreseen that only fragments - and even arbitrary parts of the national imprint of especially smaller and middle-sized countries would be incorporated, with the consequence that Anglo-American publications would, in future, dominate all levels of education, research, scholarship and public use. The response came quickly, and some will say inadequately, from the European Union on September 30, 2005, with the communication called *I2010: Digital Libraries* (The website [http://ec.europa.eu/information\\_society/europe/i2010/index\\_en.htm](http://ec.europa.eu/information_society/europe/i2010/index_en.htm) has a good overview of the policy actions and documents within the field). This was followed up by an extensive hearing process within the library, archive, museum and cultural sectors of Europe, being finalised by the *Recommendation on the digitisation and online accessibility of cultural material and digital preservation* by the Commission on August 24, 2006, agreed by the Ministers of Culture in November 2006. This statement

provides the framework for digitisation policy actions of the European Commission in the years to come, including the *Europeana* project based on the existing service *TEL - The European Library*, a portal introduced by the Conference of European National Librarians some years earlier.

According to the American Council of Learned Societies (ACLS) Report on Cyberinfrastructure, "Our Cultural Commonwealth" (2006), Google Book™ represent only about a third of the books held in research libraries — and there are many forms other than books in which the cultural record is preserved, not to mention the fact that many books are not held by public research libraries. In public and non-profit digitisation efforts, as these are proposed, priority must be placed on those collections that commerce is unlikely to fund.

#### Seamless access to data

One major objective is to provide seamless access to data across repositories, nations and research purposes. Data generated for one purpose should be open for use in many ways. They should consequently be created, described, and preserved in ways that facilitate use for a variety of purposes. The use of standards is crucial to the sharing of data. Yet the culture of standards has shown to be weak in the Social Sciences and Humanities.

To encourage data sharing and interoperability across communities and software systems, standardisation of metadata and data are needed and should be encouraged. For Social Science data, important initiatives and efforts are underway at an international level, most notably the Data Documentation Initiative (DDI) undertaken by an international alliance of data producers, curators and disseminators.. Another initiative is the Statistical Data and Metadata Exchange (SDMX) set up by data producers at the international level, like the International Monetary Fund (IMF), Eurostat and the European Central Bank (ECB) and which is used by several European national statistical organisations.

However, in the vast majority of cases where



data are put into online information systems, the resources cannot be searched or analyzed across different platforms. Each resource or data collection residing in its own silo – few with adequate curation or long-term preservation procedures. Thus full advantage cannot be taken of new technical possibilities. Quite often the problems from the analogue world are being replicated and the result is that researchers are unable to assemble physically dispersed information in an easy way.

Within the Social Sciences the majority of existing European data are by design not comparative. However, data collected at a national level might still be of value for comparative research if they are properly documented and harmonized. An example of this type of post hoc documentation and harmonisation of data is the work carried out by the Luxembourg Income Study or the efforts undertaken by ICORE (International Committee for Research into Elections and Representative Democracy) regarding election studies.

Documentation and harmonisation of data for comparative research is, however, a time-consuming and expensive activity that seldom will be funded by individual research projects. It should be addressed at the European level and funded as an infrastructure activity.

To make sure that all important requirements are met, the European research community must play an active part in the development of standards. It is even more important to encourage and facilitate the use of metadata standards for documenting and publishing the existing inventories of research data available from national data archives and research institutes across Europe. This is a major undertaking that can only happen if the relevant incentives, tools, training and support networks are put in place.

#### Interoperability

One of the major challenges in the building of new European research infrastructures lies in the interoperability between different domains of research. Currently, research infrastructures are not only separated across

national borders, but also by disciplinary fields and, above all, by a long tradition of isolating the Humanities from the Social Sciences and vice versa.

It is obvious that technological and semantic interoperability are vital in enabling and creating pan-European (and beyond) interdisciplinary access to the services provided by research infrastructures.

Likewise, another invisible border can be identified which results from the diversity of languages within the European Research Area. Because of language barriers many digitized objects, data bases, and so on, are only available for researchers using the specific language in which they have been captured. Researchers outside that language community cannot access important national data or nationally available digitalized objects. Thus, overcoming language barriers is an essential need both for the Social Sciences and for the Humanities.

One of the main tasks of the new European research infrastructure will be to leave these old distinctions and divisions behind and create research tools and instruments that can prevent fragmentation, and that can be readily accessed and used both by the Social Sciences and Humanities, irrespective of language differences.

#### Central data access/location services

In the majority of European countries data are preserved and made available for research communities by national repositories. It is of crucial importance that these data resources can be found and accessed through central web-based cataloguing and location services. This does not imply that data as such should be centralised in a “European Data Archive” – only that similarly structured data and metadata are made available through a virtual and central access point.

Among the functions that should be supported by central location and access service are: a) a registration service allowing existing and new resource providers to register their data with the system; b) a resource location service, allowing researchers and other end users to find data



easily; c) functionality that facilitates easy identification of comparable data across datasets and sites; and d) functionality that provides multi-lingual resource-location services.

Technologies meeting a subset of these requirements have been developed by EC-funded projects, like MADIERA and LIMBER. The Council of European Social Science Data Archives (CESSDA) is now active in developing a European data location and access service, building partly on the technologies coming out of these projects.

#### Harmonization of data access policies

One major obstacle for access to empirical data in Europe is produced by the multitude of data access policies and regulations implemented by national governments. To make data available easier for cross national research a mapping of data resources in various countries should be made, followed by the establishment of harmonised access regulations.

In an ideal world, a European-wide agreement for data that provides a central registration and authentication service for scientific data users could be envisaged based on state-of-the-art PKI-technology. A service like this could also function as a source of knowledge about national access regulation, and have as one of the tasks to promote open access and data sharing.

### Long-Term Strategic Goals

Having identified major current and future needs for research infrastructures the next task at hand lies in the explication of major long-term goals and strategies that should become the targets or attractors for the process of European construction for the process of construction of European research infrastructures in the area of the Social Sciences and Humanities. In essence, three major goals can be identified that address the six different infrastructure needs. These goals can be formulated as:

- increasing the availability of quality comparative data and enhancing data modeling;

- the development of data integration and language tools;
- coordination of and facilitating access to research resources and materials through integrated systems and processes enabling;
- the digitization, enhancement and repurposing of key analogue data resources.

The first of these strategic goals is linked to the need for European-wide data; the second encompasses the interoperability of data and languages while the third is directed towards the harmonization of data access policies, as well as creating greater interoperability between existing and emerging research infrastructures across the Social Sciences and the Humanities in general. The last, is particularly relevant to the Humanities, recognising the special and large scale problem which exists within these disciplines which impacts both on resource discovery and access, as well as limiting methodological development.

#### European Comparative Data and Modelling

More concretely, the first strategic goal lies in the specification of best practice standards for European data collections, both for survey and for panel research. Here, the European Social Survey has already established itself as a European data infrastructure for surveys which has been explicitly founded on the premises of being the European standard in terms of data reliability, data quality and data comparability.

A similar best practice standard has also to be created for panel data where currently SHARE offers the prospect of becoming the European model for highly reliable and comparative panel data.

Finally, a substantial number of current comparative data programs like International Social Survey Programme (ISSP), European Value Studies (EVS), or European Election Study (EES) for instance should be linked to a model in the same way as ESS or SHARE so that European data collection proceeds, as a long-term strategic goal, in a similar high quality fashion with a growing stock of best



practice tools and instruments necessary to conduct European surveys or panels.

#### Data Integration and Language Tools

Turning to the second strategic goal one can see two major tasks that are largely independent from each other. The first task lies in the field of data-integration where the long-term vision is to grant access to researchers from the Social Sciences and the Humanities across Europe to all relevant nationally collected data as well as to all comparative data collections. The current model for this vision can be seen in the dissemination policies of the ESS which becomes available free of charge to the entire European research community as soon as the necessary data integration and data cleaning processes are completed. This is augmented by the proposed methodology for the DARIAH data collection and preservation methods for the Humanities that have been already accepted in principle by many EU countries.

Likewise, the second long-term task lies in the proliferation of tools and instruments, likes what CLARIN offers, which help to reduce the language barriers which currently inhibit the access to many national collections or to understand the available texts or documents linked.

#### Coordination and Enabling

The third long-term target is probably the most difficult to accomplish and it lies in the enabling and in the overall coordination, very broadly speaking, of the interoperability of research infrastructures across national borders, across scientific fields as well as across language barriers in Europe. CESSDA and DARIAH represent initiatives in this direction.

Difficult as it may sound, a substantial number of steps must be taken within the next years towards the third long-term goal, especially because the new information and communication technologies, if implemented in an intelligent pan-European manner, would allow for a giant leap towards harmonisation, standardisation and accessibility both of data and document collections across Europe.

If the new information and communication technologies are implemented within the Humanities and Social Sciences in an optimal way, a giant step towards harmonisation, standardisation and accessibility both of data and documents across Europe can be realised.

The starting point must be with cooperation between the present SSH RWG Roadmap projects.



## 4 TOWARDS AN INFRASTRUCTURAL SYSTEM



“The overall objective of the ‘Research Infrastructures’ part of the FP7 Capacities programme is to optimize the use and development of the best research infrastructures existing in Europe” ([http://cordis.europa.eu/fp7/capacities/research-infrastructures\\_en.html](http://cordis.europa.eu/fp7/capacities/research-infrastructures_en.html)). ESFRI’s previous choice of recommended infrastructure projects, as well the recommended Task Force activity, reflects this goal and thus major infrastructural needs.

The RWG recognises that research activities within the Humanities and Social Sciences in Europe and beyond are covering broad and diverse fields with multiple entry points for researchers. There is no doubt that both fields need access to high quality infrastructural services. To follow up the researchers’ need one has to focus on both the capacity and quality of the research infrastructures where one recognises the complexity and scale of the disciplines. It has to be responsive to emerging opportunities and new ways of working.

To try to meet some of the stated needs, the RWG has discussed with the coordinators and representatives of the present ESFRI PPP projects a strategy which ensures that we will have a stronger research infrastructure for the SSH field than what can be accomplished by the five projects working alone. As part of this strategy the RWG recommends that ESFRI explores ways to support a system of better cooperation between the PPP projects with the aim of developing and incorporating new forms of collaborative activity designed to enhance the research infrastructures.

It is proposed to start with a pilot study with the overall aim of developing a collaborative system such as that described above, and to test whether such a system provides the basis for a coherent pan-European research infrastructure on a larger scale for the Humanities and Social Sciences.

In the 2006 ESFRI Roadmap Report six projects within the Humanities and the Social Sciences were brought forward as mature and pan-European. The six were:

- Council of European Social Science Data Archives (CESSDA)
- Common Language Resources and Technology Initiative (CLARIN)
- Digital Research Infrastructure for the Arts and Humanities (DARIAH)
- European Research Observatory for the Humanities and Social Sciences (EROHS)
- European Social Survey (ESS)
- Survey of Health, Ageing and Retirement in Europe (SHARE)

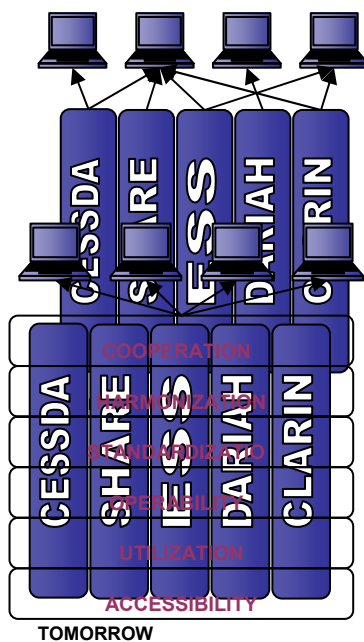


Five of the projects applied and received EC funding (FP7) for a Preparatory Phase Project (PPP) to establish the legal, organisational and financial framework for the construction and operation of the new or updated infrastructure. EROHS did not apply.

The five PPP Projects are all based on digital technology, digital tools, and digital information, in one way or another.

## Cooperation

The current situation is one in which the existing infrastructural services more or less are working on their own with a number of access systems, different access policies, data and metadata standards, and formats.



In order to obtain more time-effective, cost-effective and richer research infrastructures a number of different services should be included in a system based on close cooperation.

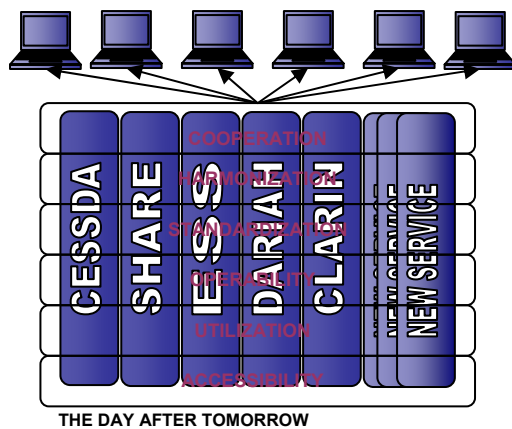
A collaborating system would provide a better breeding ground for the development and transfer of ideas, standards, experiences and skills between the cooperating facilities. The starting point should be the five SSH-projects from the 2006 ESFRI Roadmap, now funded by the European Commission for a Preparatory Phase Project (PPP).

The fact that the PPPs are diverse and provide different services in different ways provides opportunities to benefit from obvious comparative advantages. The diversity of the projects could prove to be one of the greatest qualities of the cooperative system. Knowledge transfer is vital to enable innovative new standards, tools, services, practises and policies to be developed.

A system that not only provides research communities with better services in a variety of ways, but also has a proven ability to cross-fertilise and provide advantages and cost-benefits to the participating facilities will have a better chance to succeed in the future. If such a system of



infrastructures is functioning well, it will also be attractive for other existing and emerging services to join it at a later stage.



A collaborative infrastructural system would basically require:

- commitments from the participating facilities;
- linkages to other key players both inside and outside Europe;
- funding for joint coordination and developmental activities.

This does not imply the establishment of a single infrastructure that would solve all current problems of access to relevant data, metadata and other infrastructural services. It is, however, of crucial importance that the various communities of data producers, data providers, tool developers and suppliers should have a single focal point that provides a platform for cooperation and for addressing the problems and challenges in creating and running infrastructural services.

It is the RWG's conviction that both the Humanities and the Social Sciences and the research infrastructures themselves will benefit from cooperation between the research infrastructures. How to best organise this should be a central task of the proposed pilot study.

### **Initial Phase**

In the first phase, and as a starting point, cooperation should be based on and emerge from the challenges and tasks that the existing PPPs will probably face and deal with as part of the Preparatory Phase. Later on, when a proper cooperation is established between these projects and necessary experience is gained, the cooperation should be opened up for other research infrastructures and research networks to align themselves.

The PPPs have reported that they have a number of similar or overlapping work packages. The sharing of experiences and knowledge will therefore be of mutual advantage for the projects. The need for collaboration through the Preparatory Phase and at a later stage has been expressed by the PPP projects.



## ***Need for Elaboration***

A first step on the way to cooperation would be to prepare a report identifying possible areas of overlap, potential for co-working and cross-fertilisation, and the gaps between the scientific needs, services offered, and how possible future collaborating system should be organised. Such a report should describe the nature of the cooperation and how to proceed in order derive maximum cooperation across research infrastructures and minimize the gaps to the benefit of the researcher communities.

The analysis should put special emphasis on developing specific recommendations concerning:

- coordination and cooperation between infrastructures;
- accessibility to data held in archives and other repositories;
- interoperability and harmonization of infrastructures and information;
- standardization of tools and digitized materials;
- secure utilization of data and documentation by the research communities;
- upholding of legal rights of both data subjects and data owners.

Given that there now are five SSH research infrastructure projects that agree upon the need for cooperation and also would like to work together gives the opportunity for further development of an SSH research infrastructure. The proposal is to start with an independent report, for which separate external funding will be required.

## ***A Cooperation Task Force***

A well functioning collaboration has to be developed on the basis of the competence the institutions involved in the infrastructures have. They all are or will be based at institutions that are well-developed and highly skilled. Currently they are providing services to research communities. It is however, necessary to establish a new body that will help organise and support the coordination of collaborating activities, facilitate knowledge transfer, encourage the incorporation of new infrastructures, provide services and advice to existing and emerging infrastructures and research networks, and maintain and secure continuation of the cooperation. The SSH RWG's proposal<sup>2</sup> is thus to establish a new Task Force to take forward and develop further the work of the RWG.

Such a Task Force will need to be based on the active support and participation from researchers and academic communities with a broad scientific background from the fields that should be serviced by the Social Sciences and Humanities Research Infrastructures. The SSH RWG

---

<sup>2</sup> Jane Ohlmeyer, Ireland, and Gert Wagner, Germany, did not support the proposal to form a Task Force

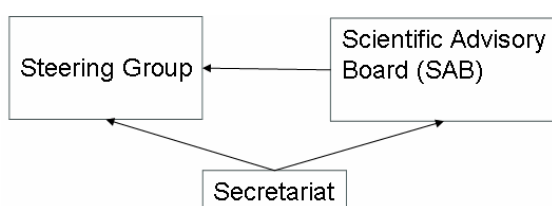


proposes that the Task Force consists of two separate bodies:

- a Steering Group;
- a Scientific Advisory Board (SAB) that provides advice to the Steering Group.

The members of both bodies need to cover a broad scientific competence. They should be appointed on the basis of consultations with European Science Foundation, ESFRI and the countries participating in the establishment and infrastructures that will participate actively in the cooperation must be represented in at least one of the two bodies. The rules for appointments need, however, to be discussed in more detail.

The Task Force will need a supporting secretariat to coordinate and facilitate the work of the Task Force<sup>5</sup>. The secretariat should serve both the Steering Group and SAB.



Task Force Model

In order to avoid conflict of interests, achieve mutual trust among the infrastructures and towards the research communities, and to maintain a well-functioning balance between the affiliating partners, it is recommended to locate the secretariat for the Task Force in a 'neutral' setting outside the infrastructures themselves.

Although it is the membership of the Task Force groups that must have the necessary research infrastructure competence, it is essential that the secretariat has enough competence to accomplish its duties. This competence should probably include knowledge of the Social Sciences and Humanities and their demand for data and documentation, and broad overview of standards, formats, tools, etc. Knowledge about the scientific landscape of Europe and the rest of the world should also be of value. However, the tasks, functions and thus the competence needed within the secretariat should be elaborated in detail during the pilot phase.

The secretariat for the Task Force needs to have a certain critical size to be able to build up and maintain enough capacity to achieve its objectives. At this stage it is, however, not possible to be very specific on numbers, but should be left to the pilot study. It is vital that the Task Force is able to maintain an enduring operation. This requires a long-term and stable funding. Different models of funding should be also looked into as part of the pilot project. However, it is important to note that it is not likely that the existing PPPs should or could pay for coordinating activities over their present budgets. There is simply no money allocation for such an activity within the specified budgets of these projects.



## 5 ANNEXES



### Annex 1: Terms of Reference

*Last update 9 March 2007  
Terms of Reference of Road Map Working Groups (RWG's)  
after September 2006*

#### Preamble

- The ESFRI roadmap, as approved in September 2006, has been built to identify needs of the European research communities for new or major upgrades of pan-European Research Infrastructure (RI), covering all scientific areas.
- All selected projects have been identified as important to impact on the science and technology development at international level, supporting new ways of doing science in Europe and the growth of the European Research Area (ERA).
- New RI's (or major upgrades) in the Roadmap have different degrees of preparedness, but all of them need to find long-term commitments by a relevant European partnership including intergovernmental organisations.

This needs to be effectively realised as soon as technically feasible. Therefore the roadmap has been presented to all relevant actors, including Governments and the EU.

- The ESFRI roadmap is an ongoing process and the construction of the scientific landscapes has also identified a number of Emerging proposals, as well as some further Pan-European needs, to be taken into account in view of the future periodic updates.
- The European Commission, on its side, will use the list of projects in the Roadmap as a base to activate preparatory actions to facilitate their realization, during the 7th Framework Programme, and will be supporting design studies to help the maturity of emerging proposals.

#### Rationale for the RWG's

- ESFRI has decided to set-up the Roadmap Working Groups (RWG) as specific Working Groups aimed at supporting, from a scientific point of view, a coherent and strategy-led approach to European policy making on new RIs of pan-European interest.
- Specific Road Map Working RWGs have been set-up, for activities related to the Roadmap in four main areas as identified by ESFRI members. These are the Social Sciences and Humanities



(SSH), the Environmental

Sciences (ENS), the Bio and bioMedical Sciences (BMS), the Physical Sciences & Engineering (PSE)<sup>1</sup>.

- The RWGs should fulfil an ongoing strategic process to help ESFRI and its stakeholders to implement and update the Roadmap.
- It is reminded that only ESFRI is responsible for the identification of individual projects within the roadmap.

<sup>1</sup> In addition the e-IRG group is providing input to ESFRI in the e-infrastructure domain

#### **Creation and update of an RWG**

- An RWG is set-up by ESFRI and chaired by an ESFRI member, having expertise and/or qualifications in the general field of the RWG that would inspire the confidence of that community. The nomination of the chair of an

RWG is made according to the rules of procedure of ESFRI.

- The duration and composition of the RWG, its field of activity, and its specific terms of reference are indicated by ESFRI. ESFRI will also update the composition, field and terms of reference, as necessary.
- It is essential that RWGs represent the whole range of topics to be discussed. On this basis, the members of the Group are proposed and finally nominated after discussion with the chair through the national delegations (with CV)<sup>2</sup>. They should have high scientific and/or managerial qualification and integrity, capable to integrate his/her contribution in developing a strategic and independent science-policy advice. If the balance of the nominations is not correct the RWG chair alerts the ESFRI chair, who in turn alerts the ESFRI delegations.
- In all applicable cases, with the agreement of ESFRI, the RWG should liaise with the intergovernmental and/or international organizations already acting in the relevant fields.
- ESFRI agrees the final composition of the RWG, which is then published on the website.

#### **Activities**

- The activities of the RWGs are the following:
  - Monitor and advise ESFRI on the best procedure to support the efforts for the implementation of the Roadmap, reporting to ESFRI on the specific RIs and on strategic aspects and/or bottlenecks, e.g. suggesting proper initiatives to support negotiations between the interested parties and/or overcome the lack of appropriate institutional arrangements;
  - Follow the general development of the EU and Global RI landscape in the field, advise ESFRI on the best procedures to stimulate the maturity of emerging or missing proposals and evaluate new (limited amount of) proposals forwarded by ESFRI delegations.
- The process of evaluation of existing and new projects must be transparent in every aspect of its definition, in order that all stakeholders can be confident in the final recommendations;
- The RWG chair is responsible for the timetable and good organisation of the meetings, for which



he/she may be helped by an EC official assigned to each Group<sup>3</sup>.

- RGW reports are regularly presented to ESFRI and published in the ESFRI web site.

<sup>2</sup> Member State could decide whether they wish to be represented in every RWG provided that the group is balanced;

<sup>3</sup> Other EC official(s) who are specialists in the field might attend the meetings as observers;

#### **Method of working**

- Meetings of the RWG will be held in closed-sessions; all information exchanged and prepared within is meant for internal use only, unless explicitly stated.

- Members of the RWG must declare any conflict of interest whenever it occurs, in relation of specific projects being discussed; such conflicts will not necessarily bar them from participating in the Group.

- Each Group should discuss the need and best way to protect the confidentiality of their deliberations, to allow for open discussions within the group. In addition, while keeping internal discussions confidential, the

method of working should allow the group to communicate with the scientific community in the most appropriate way, in order to gather all the necessary information and assure the transparency of the process. Every different opinion should be recorded in the reports to ESFRI.

- The RWG may seek independent scientific, technical or socio-economic advice, making use, as necessary, of existing bodies and/or specific experts e.g. through workshop or time-limited expert groups to deepen the discussion in specific fields.

- The RWG should not become the expression of any specific lobby-group supporting or opposing a specific proposal.

- The RWG should not generate itself projects but should advise ESFRI on the best process to stimulate important proposals which are missing in their fields, and evaluate the proposals emerging from this process.

- The Groups should meet and consult as necessary and in accordance with the needs of the timetables set down by ESFRI.

- The Chairs of the RWG should report to ESFRI whenever required by ESFRI. Specific evaluations of projects proposed for the Roadmap and specific suggestions for strategic initiatives should be presented in written reports, detailing procedures followed and experts used.

#### **Resources, time scale, deliverables and review**

- Resources to cover travel expenses of RWG members will be covered by each delegation.

- Chairs of the RWG must be able to provide their own secretarial support.

- The Roadmap and RWG deliberations are an ongoing process, and ESFRI indicates the required timescales for specific deliberations.

- General information on the RWG activities and the RWG reports should normally be circulated through the ESFRI Secretariat.

- The work, rationale and composition of each Group will be reviewed by ESFRI on a yearly basis, and after approving each edition of the Roadmap.



## ***Annex 2: Procedures and added criteria***

ESFRI added a set of criteria to the rules for review of project proposals submitted from ESFRI to the Roadmap Working Groups. These are as follows:

- Evidence that the management of the infrastructure will ensure open access to all interested researchers, based on quality of the users proposals.
- Evidence that the infrastructure is either new or proposing a major upgrade, and that this is fully justified by the quality and potential increased service to the scientific community.
- That the proposal is not only seeking a "EU label" to become more visible based on this aspect in place of its quality and pan-European value.
- Interconnection or possible links with several proposals (some already on the Roadmap) and across disciplines, leading to the construction of some sort of complex "infrastructural systems" (the RWG's should try to understand how they fit in the scientific landscapes).
- Links with national roadmaps or other vision & strategy documents
- If necessary, consider the proposal in the context of the overall landscape in order for the RWG to define the priorities between the various proposals



### Annex 3: Template use for reviewing the proposal

*Assessment of proposed research infrastructures (RI)  
for the update of the ESFRI Roadmap 2008 in the Social Sciences and  
Humanities.*

Name of proposal:

---

Name of evaluator:

---

Scientific Case	Comments
Does the proposed RI offer an important service to the future needs of users?	
Are the users identified?	
Is the pan-European (or Global) value clearly demonstrated?	
Will the proposed RI fit into the existing and future landscape of research and RIs?	
Does the proposed RI contribute to pan-European (or Global) harmonization and standardization?	
Will the proposed RI ensure open access to the research community?	

Concept Case	Comments
Is the necessary scientific and technological expertise identified?	
Is the concept technologically feasible?	
Is the requirement for e-Infrastructure sufficiently described?	
Are cost estimates feasible?	
Will the proposed RI ensure open access to the research community?	
Is the proposed RI only seeking an "EU label" in order to become more visible?	



## Annex 4: Assessment of the Received Proposal

Assessment of proposed research infrastructures (RI) for the update of the ESFRI Roadmap 2008 in the Social Sciences and Humanities.

Name of proposal: European Election Studies (EES)

Name of evaluator: RWG SSH (Dominique Joye, Karl H. Müller, Jan Karel Koppen) (meeting RWG: January 11, 2008)

### Scientific Case

#### Does the proposed RI offer an important service to the future needs of users?

This RI is an infrastructure for research into citizenship, political participation and electoral democracy in the European Union. It is used to address fundamental scientific questions regarding representation and democracy in the European Union. An upgrade is under way and will improve in consistency in scope and data quality.

#### Are the users identified?

The user community extends to every country in the EU, in the first place to researchers in European studies, political economy, political science and social psychologists, but also to users outside the scholarly community.

#### Is the pan-European (or Global) value clearly demonstrated?

This RI is a unique database in continuous existence since 1979. 24 European countries have contributed financially to the 2004 study.

#### Will the proposed RI fit into the existing and future landscape of research and RIs?

The RI has links to other databases and provides tools for innovative applications in cross-national, cross-regional and cross-industry research.. However, no clear connections have been made with other SSH roadmap RI's like CESSDA and DARIAH. There is no evidence of linking to another international project on election studies, called CSES (database of national election studies).

Preferably, the EES should propose a large integrated RI including CSES and possibly other large survey and panel studies to ESFRI and with established links with RI's of the SSH roadmap.

#### Does the proposed RI contribute to pan-European (or Global) harmonization and standardization?

There is no evidence that the RI promotes these, apart from products and tool-kits produced in several projects through the RI.

#### Will the proposed RI ensure open access to the research community?

Existing data have been archived at data archives in the USA, UK, DE. A user interface is under construction. Even if it is in the philosophy of the project, this is a very important point to consider.

#### Is the proposed RI only seeking an "EU label" in order to become more visible?

The EES-proposal clearly states that it is not expecting to receive funding from ESFRI at the time. Instead, EES wants recognition as a mature RI, by being placed on the European roadmap.



**Concept Case****Is the necessary scientific and technological expertise identified?**

The proposal does not give detailed information on PI, research and support groups concerning EES. The contactperson, however, is a top class researcher in the field. The network presented on the web size is also impressive.

**Is the concept technologically feasible?**

The RI is already up and running, going through an update, financed by the EC (until 2012)

**Is the requirement for e-Infrastructure sufficiently described?**

No special requirements are requested

**Are cost estimates feasible?**

Total preparation costs are estimated at M€3.3, of which M€2.4 is already committed by the EC and the rest by universities. Additional costs (M€2.6) are not requested from the ESFRI community.

In the case of an integrated project with, among other, CSES, a new estimation has to be done.

**Will the proposed RI ensure open access to the research community?**

## Annex 5: Composition of Roadmap Working Group

<b>Participating Members</b>	
Bjørn HENRICHSEN	Chair
Karl MÜLLER	Austria
Bogdan VAN DONINCK	Belgium
Erland Kolding NIELSEN	Denmark
Maria THEOFILATOU	European Commission
Rajola LEA	Finland
Gert WAGNER	Germany
Victoria TSOUKALA	Greece
Istvan KENESEI	Hungary
Jane OHLMEYER	Ireland
Sebastiano BAGNARA	Italy
José MENDES	Portugal
Jerca VODUŠEK STARIČ	Slovenia
Louise MCNALLY	Spain
Janne JONSSON	Sweden
Dominique JOYE	Switzerland
Jan-Karel KOPPEN	The Netherlands
Ozlem SARI	Turkey
Kevin SCHÜRER	United Kingdom
<b>In addition the following were appointed</b>	
Rein RAUD	Estonia
Jean-Emilie TOSELLO-BANCAL	France
Brian WARRINGTON	Malta
Vasilica CIUCĂ	Romania
<b>Secretary</b>	
Dag KIBERG	



## Annex 6: Roadmap Working Group's meetings

For the update of the second Roadmap the SSH Roadmap Working Group has held three meetings. In addition a meeting was held between the coordinators of the five SSH Preparatory Phase Projects and the Chair and Secretary of the Roadmap Working Group. In the following an overview of the activities is given.

<b>SSH RWG Meetings</b>	
<i>Date</i>	<i>Venue</i>
October 25, 2007	British Academy, London, UK
January 11, 2008	NH Hotel, Schipol Airport, The Netherlands
April 24, 2008	Radisson SAS Hotel Norge, Bergen, Norway
<b>Special invitations<sup>3</sup></b>	
<i>Name</i>	<i>Affiliation</i>
Tamás VÁRADI	CLARIN (October 25, 2007)
Roger JOWELL	ESS (October 25, 2007)
Hendrik JÜRGES	SHARE (October 25, 2007)
Sheila ANDERSON	DARIAH (October 25, 2007)
Cees VAN DER EIJK	European Election Study (January 11, 2008)
Algis KRUPAVICIUS	European Election Study (January 11, 2008)
<b>Meeting between PPP Project coordinators and RWG Chair &amp; Secretary</b>	
<i>Date</i>	<i>Venue</i>
February 19, 2008	Royal Institute of British Architects, London, UK
<i>Name</i>	<i>Affiliation</i>
Kevin SCHÜRER	CESSDA
Steven KRAUWER	CLARIN
Peter DOORN	DARIAH
Roger JOWELL	ESS
Hendrik JÜRGES	SHARE
Bjørn HENRICHSEN	SSH RWG – Chair
Dag KIBERG	SSH RWG - Secretary

<sup>3</sup> To the meeting in October 2007 coordinators for the SSH PPP projects were invited to update the RWG about status and progress of the projects. Kevin SCHÜRER, the coordinator of CESSDA PPP, is a member of the RWG and is hence not included in the "Special invitations" list.



### **Annex 7: Declarations of Conflicts of Interest**

Following the Terms of References (see Annex 1) the members of the SSH Roadmap Working Group declared the following potential conflicts of interest, as listed below. These potential conflicts did not necessarily bar them from participating in the discussions of the Working Group.

	ESS	SHARE	CESSDA	CLARIN	DARIAH	EROHS	EES
<b>Bjørn HENRICHSEN</b>	X		X			X	
<b>Louise MCNALLY</b>	X			X			
<b>Dominique JOYE</b>	X	X	X				
<b>Karl MÜLLER</b>	X		X				
<b>Jane OHLMEYER</b>					X		X
<b>Kevin SCHÜRER</b>			X		X	X	X
<b>Gert WAGNER</b>						X	
<b>Bogdan VAN DONINCK</b>			X			X	
<b>Jerca VODUŠEK STARIČ</b>					X		
<b>Victoria TSOUKALA</b>				X	X		
<b>Istvan KENESEI</b>				X			



## **Annex 8: Proposal Received**

To be sent to the ESFRI Executive Board for consideration

Descriptive title, and information on the ESFRI delegation submitting the proposal (or one of the member of EIROForum)

European Election Study (EES)

Research Infrastructure upgrade proposal submitted by Lithuania

2. Synthesis description of the new RI (or major upgrade) and S&T fields involved at Pan EU level in its use. Add links to relevant data/web pages (half page max)

This RI is an infrastructure for research into citizenship, political participation, and electoral democracy in the European Union that has had a continuous existence since 1979 and has prepared election studies at the time of all elections to the European Parliament (EP). More details about the European Election Studies (EES) can be found on the project's website: [www.europeanelectionstudies.net](http://www.europeanelectionstudies.net). The infrastructure, which is currently undergoing a major upgrade funded by an FP7 grant from the European Commission, provides cross-national empirical data relevant to the study of public opinion formation and democracy at the European level. The data include content analyses of relevant party manifestos and of media coverage, along with surveys of political attitudes and behaviour among citizens' and candidates for elections in all member states at the time of EP elections. Past data generated by the infrastructure are publicly available through European and North American data archives and on the EES website. These data have been extensively analyzed in 17 books and more than 100 academic papers by political scientists, communication scholars, international relations specialists, political economists, and sociologists (see an incomplete list under "Publications" on the EES website). The upgrade that is now under way will improve consistency in scope and data quality across time points and across the 27 EU member states. It will also improve the access to this infrastructure and involvement of our user community, as explained below.

3. Science case: scientific area(s) and potential and/or explicit users, how the new research infrastructure will fit into the existing and future landscape of Research and of existing RI's, at EU and World level (one page max, links to relevant documents, references).

Data collected by the EES since 1979 have been used to address fundamental scientific questions regarding representation and democracy in the European Union and its member states. Elections to the European Parliament (EP) constitute the top tier of a 'multi-level system of governance'. The role of voters, parties and party leaders in this system is extremely complex because behaviour of relevant actors in EP elections is shaped by national and local politics. In turn, political representation and democracy in the member states is affected by the existence of a supranational tier of governance. The extent to which voters are able to navigate this complex system is an important empirical and normative question with implications for EU and national democratic governance.

At the same time, the new structures, and the linkages between old and new structures, also provide unprecedented opportunities for fundamental research on the mainsprings of electoral behaviour regardless of locale. Enduring questions about the role of electoral systems, party systems, and other institutional arrangements can be addressed with the aid of multi-level research



designs that take advantage of the simultaneous occurrence of elections in countries with different characteristics in regard to institutions, parties, media markets, and the like. The insights that scholars have gained from the study of electoral behaviour at the time of EP elections has been transformative in a number of ways (see the bibliography on the EES website). Moreover, the series of election studies now extends for long enough as to be useful in studies of electoral dynamics over time, as well as in the study of comparative statics. For all these reasons the RI enhances the attractiveness of research in Europe and the centrality of European research to the community of scholars worldwide.

Our user community extends to every country of the EU and funds for the 2004 study were raised in 24 EU countries (see [www.europeanelectionstudies.net](http://www.europeanelectionstudies.net)).

Hundreds of scholars are currently using the European Election Studies in their research, and the potential user community for this infrastructure is even larger, as more member states have joined the EU and the European Parliament has increasingly become one of the key institutions in the Union. In a typical political science department with about ten academic staff at a European university, it would be extraordinary for there to be less than one focusing on the study of voters, parties, and elections; and the more usual number would be two or three. In addition, scholars concerned with European studies and European politics are part of our user community, as are communications scientists, political sociologists and political economists. Social psychologists also use EES data to study the interaction between individual behaviour, group behaviour, and variety in (cultural) context. Two professional journals are devoted uniquely to electoral studies, and studies of EP elections figure prominently in their pages as well as in the pages of more general political science, communications science, and sociology journals.

Beyond the scholarly community, policy-makers and policy advisors take a keen interest in the democratic processes of the European Union. The issue of democratic legitimacy is a key challenge facing the Union today, and our infrastructure allows scholars and practitioners alike to examine developments in public attitudes and electoral behaviour. We expect such practitioners to increasingly form a part of our user community as the infrastructure is upgraded and developed to allow easy-access by non-specialists.

4. Technical case: summary of results (technical specifications) of conceptual and/or technical design studies (half page, list references/links).

The European Election Study was carried out in all member states in the election years of 1979, 1989, 1994, and 1999, and in all members but Malta in 2004. Designs of these studies are available at the EES website. For the forthcoming 2009 elections, funding is already earmarked for conducting a study in all 27 member states through an FP7 project funded by the European Commission to enhance the provision of an infrastructure for research on electoral democracy in the European Union (PIREDEU). This project has an elaborate design component - the main component of the study - which will plan a major upgrade of the EES infrastructure. Since funding for previous studies was obtained from varied sources and with different rates of success in the different years, the study design varied considerably over the years, though less for conceptual than financial reasons (see the Past Studies page of the EES website - [www.europeanelectionstudies.net](http://www.europeanelectionstudies.net)). A complete analysis of interactions between citizens, parties and political elites at the European level requires specific components focusing on each of these objects of study and on the news media that connect them. While all these components were present in one or another previous European Election Study, inadequate funding prevented their simultaneous implementation and appropriate coordination. The upgrade now underway will allow the coordinated and simultaneous implementation of all of these components in a study of the 2009 EP election



5. e-infrastructure: what does the new RI require as far as e-infrastructure? How is it integrated with the existing EU e-infrastructure (e.g. Geant, grid, digital repositories).

The EES can be conceptually divided into an ongoing project, intending to collect data at each successive election to the European Parliament, and a data repository – the latter being the result of past data collecting enterprises. The ongoing project requires little in the way of e-infrastructures beyond the communications capabilities supplied by electronic mail and file transfer facilities, which are fundamental to a multinational research project of this kind. The EES data repository, by contrast, consists of an enormous collection of individual-level data resulting from some 70,000 interviews (with about a thousand respondents interviewed in each country at each EP election) that requires linking with data concerning electoral candidacies, party manifestos, elite behaviour as revealed by media studies, and election outcomes. This multifaceted multi-level data structure has become increasingly unwieldy as it has grown in size, and a major component of the upgrade currently under way is to supply a user interface that will provide access to these data by presenting users with ‘views’ that are tailored to their needs. Using the data-cubing and other technologies that are now in their planning stage, data archives that are repositories of EES data will be able to supply users with custom datasets for analysis and custom displays for those who do not wish to conduct their own analyses. The design of this e-infrastructure has already been funded as part of a M€2.4 FP7 grant from the European Commission.

Data from past EES studies have previously been archived at the UK Data Archive; at the ICPSR, University of Michigan; and at the Zentralarchiv fuer Empirische Sozialforschung at the University of Cologne (now GESIS).

6. Other expected socio-economic impacts: development of new technologies, effects on training, involvement of industries, local impact, other (one page, references).

The European Election Study has been highly innovative not only in linking its many separate collections of data, gathered with multiple methods, but also in developing suitable techniques for the analysis of variables with inherently different codings (because of differences in party and other systems) across various contexts of interest in a multilevel research design (Eijk, Franklin *et al.* 1996; Eijk *et al.* 2006, Eijk and Brug 2007 – see the bibliography on the EES website). The EES has developed a relevant tool kit for handling such things as the different lists of parties on the ballot in different countries. This and similar innovations have a wide range of possible applications in cross-national, cross-regional and cross-industry research on European societies, economies and polities, which could greatly advance pan-European research agendas and facilitate the replacement of theory-blind country (and country-specific) names with properly theorized variables. The EES’ emphasis on linking data on public opinion to contextualized elite stimuli also has great potential for advancing the analysis of European identity formation and related research agendas in social psychology and socio-linguistic studies. The study’s combination of media, public record, and survey data analysis has also been instrumental in creating interdisciplinary linkages and in facilitating empirically-based analysis of the performance of different media systems in assisting the development of an informed electorate. At the same time the EES has provided a state-of-the-art model for election studies throughout Europe.

Given the complexity of the data sets produced and their potential to inform debates about reforming the EU’s political system and other topical issues in democratic governance, like the impact of women’s representation or the management of multiethnic polities, the EES has served as raw material for a large number of PhD dissertations over the years, and provided an attractive training tool in advanced statistical and other social research methods in contexts like the Training Mobility Research (TMR) Network funded by the European Commission and coordinated by Hermann Schmitt over a five year period from 1998 to 2003. EES data have also been used in training courses on advanced statistical methods in the Social Sciences in the ESRC Oxford



Spring School and the Essex Summer School for Social Science Data Analysis & Collection.

7. Commitments / maturity: which States / Organizations have demonstrated interest / commitment in supporting and/or funding the proposal?...

Past EESs have been funded in part by the Commission of the European Communities (1989 and 1994), the European Parliament (1994), the British Economic and Social Research Council (1989 and 2004), the Deutsche Forschungsgemeinschaft, DFG (1994) the Deutsche Bundespresseamt, (1999), the Dutch NWO (1984, 1994, 1999, 2004), the Volkswaken Stiftung (1994), the Thyssen Stiftung (1994), various universities (in all years), and the Office of the French Prime Minister (1994). Additional funding was obtained, country by country, by individual collaborators in EES2004. The EES2009 will be funded in part by an FP7 infrastructure design grant from the European Commission amounting to 2.4 Million Euros. That grant will pay for a feasibility study (conducted in the context of the 2009 EP elections) of the infrastructure upgrade that is the real purpose of the grant. Additional funding for the EES2009 will be sought from the European Parliament and national funding agencies. We should stress that the skeleton study envisaged as a feasibility study for the infrastructure upgrade will take place even without additional funding. Thus the existence of the EES infrastructure is guaranteed up until the European Parliament elections of 2014 even without additional funding. The recognition we seek as a mature infrastructure will facilitate the solicitation of additional funds for specific research projects that can be hung onto the skeleton study and will stand or fall on their own intellectual merits. We are not at this time seeking specific undertakings of financial support in connection with attaining roadmap status.

8. Costs for construction, operation and decommissioning, indications on project financing (half page, with references/links). Give budget info in M€

Total preparatory cost: M€3.3 of which M€2.4 is already committed by the European Commission and the remainder already committed by a consortium of 15 universities and research institutions from all over Europe involved in PIREDEU. Additional costs for which funding will be sought for the 2009 EES: M€2.6.

Note that as a 'support action' the PIREDEU grant only pays 7% overheads, so consortium institutions are making a larger than customary contribution to this project.

We stress again that these funds are not requested in connection with the present proposal, which is for recognition only.

Decommissioning cost: M€0. We do not anticipate decommissioning costs for a database that can be deleted at the touch of a button.

9. Timetable for construction, operation and decommissioning (half page, with references/links) with duration and possible starting dates.

A pan-European network of collaborators and national study directors is already in place with significant experience accumulated over the past 25 years. Major funding has already been secured through an FP7 project that will run from the start of 2008 until the end of 2010. The feasibility study for the upgraded facility will be conducted in 2009-2011, starting with the collection of survey data, audio-visual recording of media materials and collection of printed documents at the time of the EP elections in 2009. The content analysis of the latter will take place in 2010 and the collection of macro-statistical data will continue until the end of 2010. After 2010 we as yet have no outside funding, but data analysis and the writing of scholarly monographs and papers will continue, funded by scholars' own resources (and those of the institutions that employ them). A



follow-up election study for the European Parliament elections of 2014 has yet to be funded. However, as an established infrastructure with a proven track record we do not expect funding difficulties so long as our status has been recognized by the ESFRI.

SPECIAL NOTE: it is the existence of the ESFRI that makes it essential for the EES to obtain recognition since an infrastructure that is not recognized will find it increasingly difficult to obtain funding now that there is a formal process for acquiring recognition.

10. Reference: Person who has submitted the proposal, and will follow up in ESFRI

Albertas ŽALYS, Director of Science and Tech. Dept, Ministry of Education and Science of the Republic of Lithuania, Z. Sierakausko str, 15, LT – 2600 Vilnius, Lithuania, Tel: +370 5 219 01 12, Fax: +370 5 219 01 00,

E-mail: [albertas.zalys@smm.lt](mailto:albertas.zalys@smm.lt)



## Annex 9: References

### American Council of Learned Societies (ACLS):

"Our Cultural Commonwealth", New York, 2006

### Berman, Francine and Henry Brady:

"Final Report: NSF SBE-CISE Workshop on Cyberinfrastructure and the Social Sciences", UCLA, 2005; available at <http://www.sdsc.edu/sbe/>

### Data Documentation Initiative (DDI):

<http://www.ddialliance.org/>

### European Central Bank (ECB):

<http://www.ecb.int/home/html/index.en.html>

### European Commission:

"Commission Recommendation on the digitisation and online accessibility of cultural material and digital preservation",

[http://ec.europa.eu/information\\_society/newsroom/cf/itemdetail.cfm?item\\_id=2782](http://ec.europa.eu/information_society/newsroom/cf/itemdetail.cfm?item_id=2782), 2006

### European Election Study (EES):

<http://www.europeanelectionstudies.net/>

### European Value Studies (EVS):

<http://www.europeanvalues.nl/>

### EUROPEANA:

<http://www.europeana.eu/>

### Eurostat:

<http://epp.eurostat.ec.europa.eu/>

### Google Scholar™:

<http://scholar.google.com/>

### i2010: Digital Libraries initiative:

[http://ec.europa.eu/information\\_society/activities/digital\\_libraries/index\\_en.htm](http://ec.europa.eu/information_society/activities/digital_libraries/index_en.htm)

### International Monetary Fund (IMF):

<http://www.imf.org/external/index.htm>

### International Social Survey Programme (ISSP):

<http://www.issp.org/>

### MADIERA-project:

<http://www.madiera.net/>

### Miller, Ken and Brian Matthews:

Journal of Digital Information, "Having the Right Connections: the LIMBER Project",

<http://jodi.tamu.edu/Articles/v01/i08/Miller/>, 2001

### Milne, Ronald :

"The Google Mass Digitization Project at Oxford", LIBER Quarterly, vol. 16: 3-4, 2006

### OECD:

"OECD Principles and Guidelines for Access to Research Data from Public Funding", 2007

### Statistical Data and Metadata Exchange (SDMX):

<http://www.sdmx.org/>



## ***Annex 10: Infrastructures put forward for the Roadmap Update***

### **CESSDA - Council of European Social Science Data Archives**

**The facility:** CESSDA is a distributed research infrastructure that provides and facilitates access for researchers to high quality data and supports their use. It promotes the acquisition, archiving and distribution of electronic data, and encourages the exchange of data. The infrastructure includes 20 social science data archives in 20 European countries. Collectively they serve over 30,000 researchers, providing access to more than 50,000 data collections per annum.

**Background:** Data are the single most important component necessary for a science-based understanding of society. In recent years, there have been significant advances in making data available for scientific use but these have not been applied consistently by different countries resulting in large national differences in data availability and in the value that is attached to data access. In recent years CESSDA has developed a number of complex tools to facilitate cross-national resource discovery and data management and access. Despite these important developments which collectively form the basis of the current CESSDA Infrastructure, much work remains to be done to transform the existing set of arrangements into a fully-functioning research infrastructure in which data resources can be easily and seamlessly located and accessed.

**What's new? Which impacts?** CESSDA already has a critical impact on the social science research community by providing access to large amounts of data. The upgrade will increase CESSDA's impact by integrating the work of members and by setting standards which will enhance the one-stop shop for data location, access, analysis and delivery. Software development will increase the quality of available data. Data from sources currently outside the CESSDA membership will also become available. CESSDA will create a more dynamic knowledge management web and will contribute to metadata initiatives. The upgrade will also improve the existing technical infrastructure; promote capacity building; support less developed and less well resourced organisations; and increase CESSDA membership.

**Timeline.** The Preparatory Phase of CESSDA runs from January 1<sup>st</sup> 2008 until December 31<sup>st</sup> 2009. The length of the construction phase will be decided during Preparatory Phase but is expected to be between 3 and 5 years, dependent on funding decisions.

#### **Estimated costs:**

Preparation costs: 4.165 M€ (2.7M€ from the Commission, the remainder from partners). The Total construction costs: ~30M€ (to be determined as part of the Preparatory Phase).

Operation costs: 3 M€/year once fully developed.

Decommissioning costs: not applicable.

**Website:** <http://www.cessda.org>



## CLARIN - Common Language Resources and Technology Infrastructure

**The facility:** CLARIN is a large-scale pan-European coordinated infrastructure effort to make language resources and technology available and useful to scholars of all disciplines, in particular the humanities and social sciences. It will overcome the present fragmented situation by harmonising structural and terminological differences, based on a Grid-type infrastructure and by using Semantic Web technology.

**Background:** The volume of written texts and spoken or audiovisual recordings is enormous, and it is growing exponentially. The sheer size of this material makes the use of computer-aided methods indispensable for many scholars in the humanities and in neighbouring areas who are concerned with language material. At present 128 institutions from 32 European countries are registered as CLARIN members.

**What's new? Which impacts?** The CLARIN Infrastructure aims to provide a comprehensive and easily accessible archive of language resources and technology, covering not only the languages of all member states, but also languages studied there and language issues related to migration. The tools and resources will be interoperable across languages and domains. They will contribute to preserving and supporting multilingual and multicultural European heritage. An operational open infrastructure of web services will introduce a new paradigm of distributed collaborative development and will allow many contributors to add new services ensuring reusability and allowing scaling up to suit individual needs. CLARIN will provide preferably existing tools and solutions and the necessary training and advice to customize the resources in order to suit the particular needs of humanities researchers. It will strengthen the European position in standardization efforts, function as a pivotal and exemplary case for international initiatives and it will help Europe to train young researchers to use and contribute to an infrastructure enabling e-Humanities.

**Timeline.** The CLARIN preparatory phase, which currently includes 32 partners from 22 countries, will last 36 months (ending in 2010), The CLARIN infrastructure should reach its full functionality based on fully operational resource and service centres few years after the end of the preparatory phase.

### Estimated costs.

Preparation costs: 4.1 M€ (2008 – 2010)

Construction: 104 M€ (2011 - 2013)

Operation: 38 M€ (2014 - 2018)

This gives an overall cost of 146.1 M€ till the end of the year 2018. Decommissioning costs: not applicable.

<http://www.clarin.eu/>



## DARIAH - Digital Research Infrastructure for the Arts and Humanities

**The facility:** DARIAH provides long-term access to and preservation of research data and digital heritage materials for the arts and humanities in Europe. DARIAH connects information users (researchers), information managers and information providers. It gives them a technical framework that enables enhanced data sharing among research communities.

**Background:** The changing nature of research practices in the arts and humanities has created a pressing need for an international digital infrastructure. At the same time, developments in information and communication technology are raising exciting new opportunities for using just such an infrastructure. The total number of participating organizations is currently 14, of which two are funding organizations.

**What's new? Which impacts?** DARIAH will contribute to the efficiency and effectiveness of arts and humanities research by:

- making sure data can be found and accessed without any need for extensive travel;
- making innovative interpretation tools available to the broader research community;
- preserving data for future analysis;
- Standardizing tools and datasets to allow for interoperability.

The work in the preparatory project preceding DARIAH addresses coordination, strategic, financial, governance, logistical, legal and technical issues, as well as management and dissemination activities to support this work. The activities foreseen vary from desk research, agenda setting meetings and forums, surveys, mailings and testing of prototype technologies. The preparatory project specifically aims to deliver:

- a business plan for the construction and operation of DARIAH;
- a consortium committing to the construction and operation of DARIAH;
- a legal document laying down the rights and obligations of different types of DARIAH partners and allowing for the inclusion of new partners;
- sufficient financial support for the construction and initial operation of DARIAH.

**Timeline:** The Preparatory phase runs from September 2008 to September 2010, during which period the funding for both the construction and operation will be secured. The construction will take place during the period September 2010 - September 2013.

### Estimated costs:

Preparation costs: 3,7M€, of which 2,5M€ EU contribution.

Total construction costs: 12M€.

Operation costs: 4 M€/year preparatory phase.

Decommissioning costs: not applicable.

**Website:** <http://www.dariah.eu>



## ESS - European Social Survey Upgrade

**The facility:** the European Social Survey is an academically driven long term pan-European instrument designed to chart and explain the interaction between Europe's changing institutions and the attitudes, beliefs and behaviour patterns of its diverse populations. The original infrastructure was set up in 2001 as a time series survey for monitoring changes in social values throughout Europe and to produce data relevant to academic debate, policy analysis, better governance, and as an important resource for the training of new researchers in comparative methods.

**Background:** The European Science Foundation initiated ESS in 1998 as an embryonic infrastructure designed to provide the framework and content of a comparative time series of data. With the twin aims of monitoring attitudinal and value changes and of improving social measurement in Europe, three biennial ESS rounds have so far been conducted and a fourth round is underway, using the most rigorous methodologies. More than 30 European countries will be covered, including the EU (apart from Malta so far), several associated and candidate countries, and Russia. The infrastructure has a complex network of management and advisory committees, representing scientists in different nations, funding councils, and technical academic specialists. Its costs so far have been shared between the EC, the ESF, and 34 national academic funding bodies. The data from ESS are made public available as soon as they are available - with no prior access to anybody.

**What's new? What impacts?** As each new survey round builds upon prior rounds, data of considerable value to the social sciences are being accumulated. These data are exploited by advanced and early-stage researchers, politicians, policy makers and the public to construct long-term accounts of change and development in the European social world. However, a large and complex time series such as the ESS requires more secure continuity of funding than it has enjoyed hitherto, to enable appropriate planning and to achieve maximum impact. New organisational, legal and financial arrangements for the upgraded Infrastructure will allow experiments into new and alternative survey methods as well as extending further the reach of ESS findings to both academic and policy communities.

**Timeline:** An upgraded ESS infrastructure will embrace at least three further rounds of data design, collection and dissemination, plus associated programmes of training, dissemination and exploitation activities over 6 years.

### Estimated costs.

Preparation costs: 2 M€, of which 1.5 M€ EU contribution

Total construction and operation costs: Combining all sources of present finance, the total annualised costs of the ESS infrastructure at present is around €6m per year. The upgrade would bring the total annual costs to around €9m per year (a 50% overall increase), i.e. an overall financial commitment of around €54m over 6 years.

Decommissioning costs: not applicable.

**Website:** <http://www.europeansocialsurvey.org/>



## SHARE –Upgrade of the Survey of Health, Aging and Retirement in Europe

**The facility:** SHARE is the upgrade into a lasting infrastructure of a multidisciplinary and cross-national panel database of micro data on health, socio-economic status and social and family networks of more than 30,000 individuals aged 50 or over. SHARE is coordinated centrally at the Mannheim Research Institute for the Economics of Aging and is accessible free of charge. It is harmonised with the U.S. Health and Retirement Study (HRS) and the English Longitudinal Study of Ageing (ELSA).

**Background:** SHARE was initially set up as a part of the Fifth Framework Programme key action on Population Ageing. Eleven European nations were included in the first SHARE wave (2004 to 2005). Four more nations were added to the second wave of data collection, which has been supported through an Integration Activity of the Sixth Framework Programme. The third wave of data collection, which specialises on the life histories, is conducted in 19 nations.

**What's new? Which impacts?** SHARE is a novel combination of inter-disciplinarily, cross-nationality, and longitudinally, each of which are necessary to understand individual and societal ageing as a process. In the context of the EU research infrastructure policy, the Seventh Framework Programme Preparatory Phase project for a major upgrade of SHARE has two dimensions: a) to generate a genuine 8-wave panel that follows individuals for up to 15 years and b) to expand to all EU Member States (plus associated Switzerland and Israel). The upgrade will therefore bring the existing SHARE prototype to the level of financial, legal, governance and technical maturity needed to achieve these objectives.

**Timeline:** the Preparatory Phase will run from January 2008 to December 2009. the upgrade will continuously run in waves every two years until 2020.

**Estimated costs.** Note that costs are roughly proportional to the number of participating countries, the number of respondents in each country, and the number of waves. Current funding is mainly from EU (almost 80%), with additional funding from U.S. National Institute of Aging and national sources. Such cost sharing is also expected in the future.

Preparation costs: 250k€ per country and wave (= 3.75M€ for 15 countries).

Total construction costs (i.e. data collection for about 2500 individuals in each country): 400k€ per country/per wave (= 6M€ for 15 countries).

Operating costs: 300k€ per year.

Decommissioning costs: not applicable.

**Website:** <http://www.share-project.org/>



Social Sciences and Humanities

# Roadmap Working Group Report 2008

## CONTACT

ESFRI Secretariat

Postal Address:

European Commission

SDME 01/66

B-1049 Brussels, Belgium

[esfri@ec.europa.eu](mailto:esfri@ec.europa.eu)

Tel +32 2 299 25 39

Fax +32 2 299 21 02

[www.cordis.europa.eu/esfri](http://www.cordis.europa.eu/esfri)

